Inferring Patterns in Network Traffic: Time Scales and Variation

Soumyo Moitra
smoitra@sei.cmu.edu
INFORMS 2014
San Francisco

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **21 OCT 2014**	2. REPORT TYPE **N/A**	3. DATES COVERED
4. TITLE AND SUBTITLE **Inferring Patterns in Network Traffic: Time Scales and Variations**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Moitra /Soumyo**		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited.

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**14**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

This material is based upon work funded and supported by SEI Line Funding under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

DM-0001793

Introduction

- A method and metrics for Situational Awareness
- SA → Monitoring trends and changes in traffic
- Analysis over time → Time series data analysis
- Metrics related to time series are key for SA

- Variations over time → Metrics for tracking
- Time windows and time scales are important to understand and interpret the metrics

Background

- Traffic patterns and variations important
  - Engineering and performance
  - SA for security – monitoring
- Important for anomaly detection
- Baselines for normalcy
- Thresholds | Inherent variations
- Alerts – Can be based on many metrics

- **Metrics based on variations in traffic**

References

Biersack, Callegari, and Matijasevic – Data Traffic Monitoring and Analysis

Box, Jenkins and Reinsel – Time Series Analysis: Forecasting and Control

Braun and Murdoch – A First Course in Statistical Programming with R

Brockwell and Davis – Time Series: Theory and Methods

Cowpertwait and Metcalfe - Introductory Time Series with R

Crovella and Krishnamurthy – Internet Measurement

Nucci and Papagiannaki – Design, Measurement and Management of Large-Scale IP Networks

Park and Willinger – Self-Similar Network Traffic & Performance Evaluation

Shumway and Stoffer - Time Series Analysis and its Applications

Method of Analysis

- Analysis of flow data to investigate this issue
- Construct an initial time series | W and b
- Establish a time slot τ (b < τ < W)
- Estimate the standard deviations within each τ
- Estimate the std. dev. of these std. dev.s **[H]**
- Compare this across varying bin sizes
- Vary time window (W)
- Compare τ-s across varying W |same bin size
- Metric can be tracked over time (successive Ws)

Variance of the Variance

1	2	3	4	5	6	7	8	9	10	
τ1					τ2					
μ1					μ2					

Variance over time (traffic load)
Burstiness ~~ Variance of the means

What about the variance of the variance?

=> Heteroscedasticity

Estimating Heteroscedasticity [H]

1	2	3	4	5	6	7	8	9	10	
τ1					τ2					
μ1					μ2					
σ1					σ2					

$\boldsymbol{\sigma}$(σ1, σ2, …) = **H**

Important to monitor **H** as well.

Data and Design

- Analysis reported here was done with public domain data
- Two time windows (8 hours each)
- Two time scales (b=4,8 minutes)
- Analysis was done with SiLK and R
- Can be done with any flow data and scripts
- One set of comparisons shown
- A particular case of heteroscedasticity

Results

Table: Heteroscedasticity Estimates

(Overall standard deviation in parentheses)

(W1 = W2 = 8 hours; b1 = 4 min, b2 = 8 min)

Time window	Time Scale = b1	Time Scale = b2
W1	13.57 MB (26.35 MB)	33.37 MB (50.10 MB)
W2	4.47 MB (10.66 MB)	7.99 MB (19.64 MB)

Conclusions

- An attack or intrusion usually implies some shift in traffic patterns
- One indicator of such shifts could be a change in the levels of heteroscedasticity
- This methodology has the potential to detect such attacks at an early stage
- Alert when H exceeds a threshold

Benefits

- This approach could detect attacks and intrusions that do not perturb the network traffic in other discernible ways
- Thus other techniques may not identify them early enough
- Early detection is important for effective mitigation

- This method also enhances SA by introducing a new metric to track traffic patterns

Future Work

Implications of changes H w. r. t. time scales?

Repeat the analysis: wide W & different networks

Predictions from attack/intrusion models <H>

Test behavior of H with data with known attacks

Thank you!

Questions/comments?